DESIGN OF A MULTI-MODAL DEEP LEARNING FRAMEWORK FOR EARLY LEUKEMIA DETECTION USING PERIPHERAL BLOOD SMEAR IMAGES AND CLINICAL DATA

BY

ONYEBUCHI DENIS IFEANYI

GOU/U22/CSC/811

COMPUTER SCIENCE AND MATHEMATICS, FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY, GODFREY OKOYE UNIVERSITY, ENUGU STATE.

JULY, 2026

DESIGN OF A MULTI-MODAL DEEP LEARNING FRAMEWORK FOR EARLY LEUKEMIA DETECTION USING PERIPHERAL BLOOD SMEAR IMAGES AND CLINICAL DATA

BY

ONYEBUCHI DENIS IFEANYI

GOU/U22/CSC/811

DEPARTMENT OF COMPUTER SCIENCE AND MATHEMATICS, FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY,

GODFREY OKOYE UNIVERSITY, ENUGU STATE.

IN PARTIAL FULFILLMENT OF THE AWARD OF BACHELOR OF SCIENCE (B.Sc) DEGREE IN COMPUTER SCIENCE.

SUPERVISOR: ENGR. CAMILLUS NJOKU

JULY, 2026

APPROVAL PAGE

This research, titled "Design of a Multi-Modal Deep Learning Framework for Early Leukemia Detection Using Peripheral Blood Smear Images and Clinical Data," has been assessed and approved by the Department of Computer Science and Mathematics Committees in the Faculty of Natural Sciences and Environmental Studies, Godfrey Okoye University, Enugu, Nigeria.

Engr. Camillus Njoku _____ _____

Supervisor Signature Date

_____ _____

External Examiner Signature Date

Dr. Frank Okebanama _____ _____

HOD, Computer Science and Signature Date

Mathematics

Prof. I. A. Ajah _____ _____

Dean, Faculty of Computing and Signature Date

Information Technology

CERTIFICATION

I certify that I completed the research included therein and that it was primarily based on my observations and investigation. Consequently, I will fully accept the University's decision if I am found to have cheated on the assignment.

_____

ONYEBUCHI DENIS IFEANYI

GOU/U22/CSC/811

DEDICATION

This project is dedicated to God Almighty, whose grace, wisdom, and strength sustained me throughout this journey. It is also dedicated to my parents and family, whose unwavering love, prayers, and sacrifices made this achievement possible.

ACKNOWLEDGEMENTS

First and foremost, I give all glory and thanks to God Almighty for the wisdom, strength, and perseverance He granted me throughout this academic journey and the completion of this project.

I sincerely appreciate my supervisor, Engr. Camillus Njoku, for the invaluable guidance, constructive feedback, and patience throughout the research process. Your mentorship greatly shaped the quality and direction of this work.

I am deeply grateful to Dr. Frank Okebanama, Head of Department, Computer Science and Mathematics, and to all the lecturers in the department for their dedication to academic excellence and for equipping me with the knowledge that made this project possible.

My heartfelt appreciation also goes to Prof. N. M. Ozofor and the entire faculty of Natural Sciences and Environmental Studies for creating a conducive learning environment.

To my parents and family, thank you for your endless love, prayers, financial support, and encouragement. This achievement belongs to you as much as it does to me.

Finally, I acknowledge all friends and colleagues who offered assistance, shared ideas, and provided moral support throughout this period. Your contributions, however small, made a significant difference.

ABSTRACT

Detection of leukaemia depends significantly on manually examining Peripheral Blood Smears (PBS), which is limited by inter-rater variation and the severe lack of specialist hematologists, especially in Nigeria and Sub-Saharan Africa. Current deep learning methods in automated leukaemia detection depend mostly on unimodal approaches that analyze PBS images alone without considering the structured clinical laboratory measurements typically considered alongside morphology by clinicians in the diagnostic process. This paper attempts to bridge these two gaps by developing a multi-modal deep learning system capable of integrating PBS images and clinical laboratory measurements. The proposed architecture includes a fine-tuned EfficientNetB3 for image feature extraction and a Multilayer Perceptron network for structured clinical data connected via a concatenation-attention fusion layer. Training and testing were carried out using the publicly available C-NMC 2019 benchmark dataset consisting of 10,661 PBS images. Synthetic feature vectors for the clinical laboratory data were obtained using class-conditional Gaussian distribution based on published clinical guidelines. On a test set of 1,600 samples, the multi-modal approach achieved 100% accuracy, precision, recall, F1 score, and AUC-ROC, performing better than state-of-the-art unimodal models in literature. The trained model was deployed as a full-stack web application using React.js and FastAPI, providing clinicians a real-time, browser-based decision support tool for leukaemia screening.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

The cancer burden is one of the biggest public health problems of the 21st century, with an estimated 19.3 million new cancer cases and almost 10 million cancer deaths worldwide in 2020 alone (Sung et al., 2021). Haematological cancers are a particularly critical group of malignancies because of their speed of disease progression, complicated nature and the challenges they present to clinical laboratory services for diagnosis. Leukaemia is one of the most diagnostically complex and clinically important haematological diseases worldwide, characterised by the uncontrolled proliferation of abnormal white blood cells of the haematopoietic system (Swerdlow et al., 2017). Leukaemia can be divided into four main types – AML, ALL, CLL and CML – which are characterised by different cell morphology, clinical course and treatment needs (WHO, 2022).

It is clinically well understood that early and accurate detection of leukaemia is crucial. For ALL and AML, the timing of diagnosis can significantly impact the therapeutic window and patient outcomes, especially for aggressive forms of the disease (Siegel et al., 2023). In Sub Saharan Africa Nigeria included, this need is further exacerbated by the critical ratio of haematologist to population and diagnostic laboratory infrastructure gaps that leave a large percentage of patients unable to be timely evaluated by a specialist (Adeloye et al., 2017). The current diagnostic gold standard for leukaemia is the microscopic examination of the Peripheral Blood Smear (PBS), which will still depend on specialist expertise and an equitable access to accurate diagnosis will continue to be a systemic challenge in resource-constrained environments.

Peripheral Blood Smear is a blood examination of a thin film of blood prepared and stained, then examined under the microscope to detect abnormalities of white blood cells, blast cell proportion, differential count, and other abnormalities (Tran et al., 2021). Although PBS analysis has been a cornerstone of the haematological diagnosis for more than 100 years, it is subject to observer variation, and may be subject to errors of fatigue, especially if performed by general practitioners rather than specialist haematologists. Although useful supportive information, automated cell counters do not replace careful morphological interpretation of the cellular abnormalities that are associated with the leukaemia subtypes. Such restrictions are the basic clinical impetus to automated, AI-aided haematological diagnostic tools (Matek et al., 2019).

Deep learning, and more specifically Convolutional Neural Networks (CNNs), has proven to have a tremendous ability in medical pattern recognition from images. Since the work of Krizhevsky et al. (2012) that first demonstrated the effectiveness of deep CNNs for visual feature extraction, a variety of architectures have been developed, such as ResNet (He et al. 2016), DenseNet (Huang et al. 2017), and EfficientNet (Tan and Le 2019), that have improved the classification accuracy of deep CNNs with progressively lower computational requirements. CNN-based methods have been good in identifying abnormal lymphoblasts, myeloblasts, and other malignant cell types from stained PBS slides to date, in haematological imaging (Matek et al., 2019; Rehman et al., 2020). In medical imaging tasks with small amounts of annotated training data, transfer learning, which involves fine-tuning a model trained on a large data set like ImageNet, has been especially successful (Esteva et al., 2017).

Figure 1.1: Sample Peripheral Blood Smear images, showing Normal cells and Leukaemic cells [Place holder - replace with actual PBS images]

Nevertheless, unimodal image-based deep learning models for leukaemia detection have a fundamental limitation: they only use visual information from the morphology of images and ignore data from structured clinical laboratory tests (CLTs) that also contains complementary diagnostic information. Clinically, haematological diagnosis is never based on PBS morphology alone and is always supported by laboratory parameters including: WBC count, differential count, blast cell percentage, haemoglobin level and platelet count (Swerdlow et al., 2017). Such an approach to diagnosis is clinically integrated and is not possible using the existing automatic systems, which is a major gap between the current AI research prototypes and the needs of a clinical deployment.

Multi-modal deep learning overcomes this challenge by developing architectures that learn from and combine data from multiple heterogeneous input sources. Ramachandram and Taylor (2017) gave a detailed overview of the multi-modal fusion strategies, showing that intermediate fusion (where the learned modality-specific representations in the branches are fused together) is a superior approach to both early and late fusion for a variety of tasks. In the 72 clinical AI studies they reviewed, Acosta et al. (2022) determined that there were 49 studies where multi-modal models performed better than their best unimodal models. Another major improvement in fusion architectures is the use of attention mechanisms proposed in Vaswani et al., 2017, which helps in dynamically weighting the contribution of each modality based upon the input, thus improving performance as well as interpretability.

In addition to the performance of models, another important, and often overlooked aspect of medical AI research is clinical deployability. The vast majority of published deep learning models for leukaemia detection are offline, research prototypes that must be programmed and run on computational hardware, and therefore are not readily available to practicing clinicians (Shen et al., 2017). Therefore, this research proposes a design and full-stack deployment of a Multi-Modal Deep Learning Framework for Early Leukaemia Detection, where the Maximum Information Nets (MINs) technique is used to integrate features from the PBS images using a fine-tuned EfficientNetB3 branch, followed by Structured Clinical Data processing using a Multilayer Perceptron (MLP) branch, and a Hybrid Concatenation-Attention fusion architecture is used to fuse them together and deployed as a complete React.js and FastAPI web application.

1.2 Statement of the Problem

The currently utilized pathway for diagnosing leukaemia at the majority of hospitals in Nigeria and Sub-Saharan Africa has some very important inefficiencies which hinder early diagnosis and poor outcomes. Manually examining PBS has been shown to be prone to inter-observer variation and error in the absence of the presence of haematology specialists in tertiary care centres (Adeloye et al., 2017). WHO (2022) found that misdiagnosis of haematological disorders in low-income settings has been shown to result in late initiation of treatment and increased mortality, especially among time-dependent leukaemia types like ALL and AML.

The current literature on the use of automated systems using deep learning algorithms for diagnosing haematological disorders from images remains largely unimodal. This means that these AI-powered diagnostic tools process PBS images independently of clinical laboratory values, which remain essential when diagnosing haematological disorders in clinical practice. As Rehman et al. (2020) demonstrated, CNN can diagnose haematological disorders from blood cells images in high accuracy. However, these systems fail to integrate the knowledge required to make an informed clinical decision like clinicians do by combining morphology and laboratory results including WBC count and blast cells percentage. The issue of being unimodal will lead to an inability of even state-of-the-art image classification models to provide reliable results in cases where the image morphology alone does not give a clear answer – something that happens very often in cases of leukaemia at its early stages or unusual presentation.

However, the lack of fully end-to-end deployable multi-modal systems for leukaemia diagnosis represents another ongoing obstacle standing in the way of successful application of AI techniques in haematology. As stated by Topol in his 2019 review on artificial intelligence in medicine, the issue of translation of AI technologies from their successful benchmark results to practical implementation in the clinical setting is a systematic issue caused by the disconnect between model development and deployment process.

1.3 Aim and Objectives of the Study

Aim

For the development, implementation, and deployment of a multi-modal deep learning framework for early diagnosis of leukaemia through integration of Peripheral Blood Smear images and clinical laboratory data using a hybrid EfficientNetB3-MLP fusion model as a web-based clinical decision support application.

Objectives

To explore the diagnostic features of images from the Peripheral Blood Smear and the clinical laboratory parameters used in detecting leukaemia and acquire and pre-process the C-NMC 2019 dataset, making the image and clinical data datasets ready for training by multi-modal deep learning algorithms.

To design and build a hybrid multi-modal deep learning model which uses a fine-tuned EfficientNetB3 image branch, a Multilayer Perceptron clinical data branch, and a concatenation-attention fusion method to detect and classify leukaemia into its multi-class subtypes (ALL, AML, CLL, CML).

To train, optimize and evaluate the proposed multi-modal system against the unimodal image-based and clinical-based systems using the classification metrics - Accuracy, Precision, Recall, F1-Score and AUC-ROC – proving the supremacy of the multi-modal system in diagnostic leukaemia.

Deploy the developed multi-modal system for the detection of leukaemia as an open-access web application through the use of React.js frontend and FastAPI backend to enable clinicians to upload images from PBS and input clinical data to get real-time diagnosis of leukaemia.

1.4 Significance of the Study

The paper contributes to the area in terms of methodology, clinical, and context. In terms of methodology, the contribution is made by advancing the use of multi-modal deep learning in haematology diagnosis through development of a hybrid fusion network model – where EfficientNetB3 feature maps and MLP feature maps are combined through attention-based fusion

technique – which is used to capture morphological and laboratory evidence towards classifying leukaemia cases. Unlike unimodal models that either consider only images or only clinical measurements, the model presented here captures a fused representation of the information, hence closer to the reasoning performed by a specialist haematologist (Acosta et al., 2022). The transfer learning from EfficientNetB3 model pre-trained on ImageNet solves the problem of scarce annotated data available in the most research settings in medical imaging area (Esteva et al., 2017).

At the clinical level, the design seeks to address an immediate and clear diagnostic requirement. Leukaemia early diagnosis leads to better patient outcomes especially in cases of ALL and AML where there is great variability in five-year survival based on the timing of diagnosis and initiation of treatment (Siegel et al., 2023). By automatically analyzing PBS image and laboratory test results, the tool provides a computational means for diagnosing leukemia and minimizing dependence on limited haematology experts, an important advantage in the context of the Nigerian healthcare setting (Adeloye et al., 2017). As far as deployment goes, the study results in developing a full-fledged publicly available web-based application which closes the translational gap recognized by Topol (2019) as the main obstacle to clinical utility of medical AI studies.

1.5 Scope of the Study

The current study considers the problem of binary classification between healthy subjects and leukaemia patients as well as leukaemia subtype classification into several categories based on Peripherial Blood Smear images and structured clinical laboratory data. The image modality data is provided solely by C-NMC 2019 Challenge (Gupta and Gupta, 2019) dataset that offers 10,661 annotated PBS microscopy images and is broadly accepted as the benchmark for haematological AI studies. The clinical data modality is presented by selected laboratory parameters including white blood cells count, percentage of lymphocytes differential, percentage of blast cells, haemoglobin concentration, platelets count and patient demographic data. Deep learning model building is performed via Python programming language by means of TensorFlow and Keras libraries, while preprocessing and model evaluation are carried out in scikit-learn library. The frontend web application is developed using React.js framework and deployed via Vercel; the inference backend API (FastAPI) is deployed using Render.

1.6 Limitations of the Study

One of the major limitations of the study is that rather than using private clinical records of Nigerian hospitals as training data, a public benchmark dataset has been used. Although the C-NMC 2019 dataset has been accepted globally as a benchmark for haematology-based AI, the patient population contained in this dataset from All India Institute of Medical Sciences (AIIMS), New Delhi, might fail to represent the full range of morphological differences of leukaemia in the clinical conditions in Nigeria due to environmental, genetic and nutrition-related factors that might impact haematological indices (Adeloye et al., 2017). Moreover, the C-NMC 2019 dataset does not contain patient-level clinical laboratory records in the dataset which required generation of statistically consistent synthetic clinical feature vectors to train the multi-modal system. Also, due to computational constraints, it was not possible to conduct exhaustive hyperparameter search. The outputs provided by the system are meant to be used only as a decision support for clinicians and the framework is not subjected to any clinical validation process.

1.7 Operational Definition of Terms

Leukaemia: Uncontrolled proliferation of abnormal white blood cells in the bone marrow and peripheral blood, with the disease classified into four types, ALL, AML, CLL and CML, according to the cell lineage and severity of the disease (Swerdlow et al., 2017).

Peripheral Blood Smear (PBS): A thin film of blood is taken from a patient and spread on a glass slide, stained with Giemsa or Wright stain and then examined using a microscope to look at the morphology of the white blood cells, the differential white blood cell count and proportion of blast cells the primary visual diagnostic tool used for haematological malignancies (Tran et al., 2021).

Multi-Modal Deep Learning: A family of deep learning architectures capable of processing heterogeneous input data from multiple modalities (e.g., images and structured tabular data) and

learning joint representations that leads to superior performance on complex diagnostic tasks (Ramachandram and Taylor, 2017).

Convolutional Neural Network (CNN) is a deep learning architecture that is used for grid structured data like images, where convolutional layers will extract features automatically in a hierarchical way of understanding the image such as edges, texture and high level morphological patterns, without manual feature engineering, which is done by Krizhevsky et al. (2012).

EfficientNetB3: Convolutional neural network (CNN) architecture from the EfficientNet family proposed by Tan and Le (2019) that systematically scales up the depth, width and input resolution of the CNN with a compound coefficient, with a state-of-the-art accuracy in image classification while using significantly less parameters than previous models like ResNet and VGG.

Transfer Learning: A deep learning approach where a model trained on a large-scale dataset, like ImageNet, is then fine-tuned for a target task of the same general type, thus transferring previously learned feature representations to the target task providing the model with excellent performance even when it has a very limited number of annotated training examples for the target task (Esteva et al., 2017).

Fusion Module: A module in a multi-modal architecture that is used to integrate feature representations from different branches of the architecture. This is a study using a hybrid concatenation-attention fusion module to dynamically weight both the image and clinical data branches (Vaswani et al., 2017).

The MultiLayer Perceptron (MLP): A class of fully connected feed forward ANNs with an input layer, one or more hidden layers with non-linear activation functions and an output layer which will be used as the clinical data branch in this study, to learn representations of features from structured laboratory parameters.

AUC-ROC: Area Under the Receiver Operating Characteristic Curve a classification performance metric not dependent on threshold which requires a model to be able to distinguish between the positive and negative class over the entire range of decision thresholds. A value of 1.0 is considered perfect classification, 0.5 is random.

Clinical Decision Support System (CDSS): A program designed to help doctors make decisions in diagnosis or treatment through combining patient data with evidence-based knowledge and delivering contextually relevant, actionable information such as prediction of the detection of leukaemias from the database of PBS images and laboratory parameters (Topol, 2019).

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

In this chapter, a thorough review of the existing literature on which the development of the multi-modal deep learning framework for early diagnosis of leukaemia based on Peripheral Blood Smear (PBS) images and structured clinical laboratory data is provided. The review is structured into five topics, including the background of leukaemia from a clinical and biological perspective; the development of deep learning techniques in the context of analysing images of blood cells; transfer learning in medical imaging; the concept of multi-modal and attention-based learning in the field of clinical diagnosis; and a detailed review of twenty relevant studies.

2.2 Theoretical Background

2.2.1 Leukaemia: Classification, Pathology and Diagnostic Landscape

Leukaemia is a cancer of the haematopoietic system in which there is uncontrolled growth of abnormal white blood cells. There are four main types of leukaemia: Acute Lymphoblastic Leukaemia (ALL) is the most common form of leukaemia in children and is characterised by cells that have not matured normally; Acute Myeloid Leukaemia (AML) is a highly aggressive adult leukaemia which occurs from immature myeloid cells; Chronic Lymphocytic Leukaemia (CLL) is the most common form of leukaemia in adults in the western world, where abnormal mature B

lymphocytes accumulate; and Chronic Myeloid Leukaemia (CML) is characterised by the Philadelphia chromosome and the fusion oncogene BCR-ABL1, which are both found in the cells of this type of leukaemia (Swerdlow et al., 2017). Both binary detection and accurate classification of subtypes is clinically necessary, as each requires a different clinical management strategy.

The typical sequence of tests performed for the diagnosis of leukaemia is a full blood count and peripheral blood smear (PBS) examination, where a thin smear of patient blood is stained with Giemsa or Wright stain and examined under a microscope to determine white blood cell morphology, blast cell proportion and differential count (Tran et al., 2021). Although the PBS is the most important initial diagnostic test for haematological malignancies, it may be subject to inter-observer variability, and fatigue-related errors, especially in resource-limited areas where trained haematologists are in limited supply (Rehman et al., 2020). These diagnostic weaknesses form the basis for the clinical motivation for automated, AI-supported haematological diagnostic tools.

2.2.2 Deep Learning for Haematological Image Analysis

Medical diagnostics, as an application of pattern recognition, has become the leading application for Convolutional Neural Networks (CNNs). The authors in LeCun, Bengio, and Hinton (2015) showed the fundamental applicability of deep CNNs to the hierarchical extraction of features from visual data, and later, the architectures such as ResNet (He et al., 2016), DenseNet (Huang et al., 2017), and EfficientNet (Tan and Le, 2019) progressively boosted classification accuracy while decreasing computational requirements. In haematological imaging, near-expert level CNN-based classification of 15 categories of white blood cell morphology from bone marrow smear images was shown (Matek et al., 2019), proving the potential of deep learning as a haematological diagnostic tool. Transfer learning using fine-tuning of pre-trained models from ImageNet trained for specific medical tasks has been demonstrated to be significantly superior to training from scratch on limited medical datasets (Esteva et al., 2017; Shin et al., 2016).

2.2.3 Multi-Modal Learning and Attention Based Fusion

Multi-modal deep learning combines diverse information from different modalities (e.g., medical images and structured clinical information) in specialized fusion architectures to generate more powerful diagnostic image representations than those from any single modality. In fact, according to Ramachandram and Taylor (2017) intermediate fusion, where separate modality-specific branches learn representations separately, and then combine them, works best across all the other fusion strategies. Acosta et al. (2022) systematically analysed 107 multi-modal medical AI studies in 14 specialties, showing that in 72% of the analyses multi-modal models performed better than the best unimodal model for the same task, and that the benefits were most dramatic in tasks involving multiple modalities and structured patient data. In particular, attention mechanisms (Vaswani et al. 2017) allow to dynamically weight the contribution of the different modalities based on the input received, which is useful for discriminative performance and interpretability, especially in the field of clinical decision support.

The datasets and benchmarks are as follows:

Annotated benchmark datasets have played a crucial role in advancing progress in AI systems for leukaemia detection. Labati, Piuri and Scotti (2011) released the ALL-IDB dataset, which consists of PBS images of confirmed ALL patients and healthy controls, as the first publicly available benchmark for automated leukaemia detection research. This C-NMC 2019 Challenge dataset (Gupta and Gupta, 2019), with more than 12,000 annotated PBS images labeled with blast cells, has since become the most widely used standard benchmark, allowing for standardised comparisons between competing approaches, and for the largest publicly available annotated PBS leukaemia image corpus. In this research, the C-NMC 2019 dataset is used as the main training and test dataset. One major challenge with all publicly accessible PBS leukaemia datasets is that none include patient-level clinical laboratory data and requires synthetic clinical feature generation for multi-modal training.

2.3 Review of Related Works

Rehman, et al., 2020, presented a ALL classification system based on VGG-16 with transfer learning technique trained on ALL-IDB1 and ALL-IDB2. Using ImageNet pre-trained weights, the model was able to reach a binary detection accuracy of 99.14% despite a limited number of annotated medical images for training. This study is considered to be successful example of transfer learning in PBS image classification. The work is unimodal; it processes a PBS image, but with no clinical laboratory data; it only detects a "binary" image; and it offers no infrastructure for deployment translating into limited clinical utility.

Anilkumar, Bindu and Jinesh (2021) provided a multi-class leukaemia classification system based on ResNet50 model with a custom augmentation pipeline on composite image dataset of 4 subtypes of leukaemia (ALL, AML, CLL and CML). The system was able to classify the different types of leukaemia with 97.8% accuracy and successfully prove the capability of deep residual networks to classify the different types of leukaemia with the help of the morphology of stained blood smears. The work is novel in showing feasibility of subtype-level classification and is purely image-based, without any data integration with clinical data or any reports of deployment in a clinical workflow.

In 2019, Gupta and Gupta held the C-NMC 2019 Challenge on ALL blast cell detection from PBS images, and published the outcome of 51 teams on a standardised 12,000+ image set. The best performing ensemble CNN models achieved up to F1-score 0.9845 in the binary ALL detection task. The challenge set the groundwork for the best results in PBS-based leukaemia detection and showed that the state of the art deep learning architectures could rival the performance of the specialist haematologist in image classification tasks. Within the scope of challenge, only binary ALL detection was tackled, with the multi-class subtype classification and clinical data fusion being left as unaddressed research directions.

Shafique and Tehsin (2018) tested their approach to the classification of acute leukaemia with two different neural network architectures: AlexNet and VGG-16, with a view to determining the effect of dropout regularisation and data augmentation on generalisation. The results of the VGG-16 model trained with the augmented training data were better. The study offers interesting empirical recommendations for regularisation in leukaemia classification for PBS. The research has no clinical data, is limited to image classification only, and no deployment of a web application.

Matek et al. (2019) created and published a dataset of 18,365 images of bone marrow smears annotated by experts, across 15 morphological categories, and trained a CNN classifier with an overall balanced accuracy of 0.80 compared with a haematologist baseline of 0.82. The study was the most rigorous published assessment of deep learning in the field of classification of haematological cell morphology, and demonstrated that CNNs can attain the performance of experts in this area. It does not include any deployment element, is specific to bone marrow smear imagery and not specific to leukaemia classification based on PBS or multi-modal clinical data.

Acosta et al. (2022) performed a systematic review of 107 artificial intelligence (AI) studies in 14 clinical specialties, such as oncology, radiology, and pathology. The review identified that multi-modal models performed best in 72% of the studies compared to the best unimodal models, with imaging with structured clinical data being the best pairing modality. It is found that the intermediate fusion is the most consistently helpful architecture class, which is attention-based. This work offers the best empirical support for the multi-modal fusion approach used in the present research.

In one study, Esteva et al. (2017) showed a deep CNN trained on 129,450 clinical dermatological images could outperform board-certified dermatologists in the classification of skin cancer, including the detection of malignant melanoma. The study is a milestone achievement of the diagnostic level of accuracy of expert-level AI models with transfer learning from ImageNet pre-trained architectures, directly setting the paradigm for the proposed image branch of EfficientNetB3 in the image segment of the EfficientNet framework.

He et al. (2016) came up with Residual Networks (ResNet), a CNN network using “skip” connections that allow one to learn very deep networks without the loss of gradient signal. ResNet has shown significant improvement in image classification in comparison with the previous architecture and has been one of the most popular backbones in medical image analysis. The

ability of ResNet50 to perform multi-class PBS detection task as shown in Anilkumar et al., 2021, has demonstrated its suitability for haematological image analysis jobs.

Tan and Le (2019) proposed EfficientNet, which is a group of CNN models that are obtained by compound scaling the network depth, width and resolution of the inputs with a single coefficient. The variant of interest, EfficientNetB3, has similar accuracy to several times as many parameters, and significantly fewer parameters and computations. Subsequent medical imaging work has shown that it is a competitive solution in clinical image applications with small amounts of training data, which makes it well suited for the medical image part of the proposed multi-modal framework, among others.

Huang et al. (2017) introduced DenseNet, a CNN structure that enables all layers to receive and send features from and to all following layers in a dense block, resulting in maximum feature reuse and gradient flow. DenseNet's performance on image classification tasks was state-of-the-art, and it showed strong performance on medical image analysis tasks with limited training data. Despite their use for haematological image classification, DenseNet proved to be more expensive than EfficientNetB3, and thus a less promising choice for the deployment-optimised backbone for the proposed framework.

Topol (2019) conducted a thorough review of the application of deep learning in clinical medicine, including AI's performance in ophthalmology, radiology, pathology, cardiology, and oncology. The study highlighted translational failure, defined as a systematic lack of integration of high-performing research AI models into clinical practice, as the key hurdle in medical AI. These findings directly inspire deployment objective of the present research which provides an implemented, publicly available, web-based application as opposed to an offline model file.

Shen et al. (2017) gave a comprehensive survey of medical image analysis applications in radiology, pathology and ophthalmology, highlighting the evolution from shallow feature based methods to end-to-end deep learning systems. The absence of user-facing applications continued to be the main challenge for real-world clinical impact, and multi-modal data integration was highlighted as one of the directions to enhance the clinical performance of AI diagnostic tools, directly impacting the architecture and deployment of the proposed framework.

A complete taxonomy of multi-modal machine learning approaches was created by Ramachandram and Taylor (2017), who classified the approaches according to the early, late, and intermediate fusion strategies and provided an overview of applications in the fields of vision, audio, and medicine. The research showed that intermediate fusion performs better than other strategies more often, especially when the modals provide complementary information. This is the case in the framework presented as it uses the complementary information from PBS images and clinical laboratory parameters.

The Transformer architecture, which used multi-head self-attention for natural language processing tasks, was developed by Vaswani et al. (2017). Since then, the attention approach has become widely used in medical imaging as it enables dynamic adjustment of weights of the modal contributions depending on each individual input. Cross-modal attention will enable the framework to discover which clinical variables are the most relevant based on the particular image's morphology.

Labati, Piuri, and Scotti (2011) presented ALL-IDB dataset, which consists of two sets cular images of annotated PBS microscopy images of ALL patients and healthy people) as the first publicly available benchmark specifically intended for automatic leukemia diagnosis. This benchmark allowed comparing different early computational approaches to the blast cells detection and is considered an important historical landmark in the field of leukemia AI diagnostics. Its main drawback lies in the lack of the corresponding clinical laboratory information and in its smaller size compared to the subsequent C-NMC 2019 dataset.

Sung et al. (2021) have recently reported about GLOBOCAN 2020 statistics on global cancer, stating that there were 19.3 million newly diagnosed cancer cases and 9.96 million cancer deaths around the world in 2020. Haematological cancers were found to be among the most rapidly growing types of cancers in low- and middle-income countries where insufficient diagnostic facilities can cope with the increasing number of diagnoses thus proving the necessity of addressing the

diagnostic problem within the suggested framework.

Shin et al. (2016) carried out an empirical study of various transfer learning techniques applied to medical image classification tasks, proving that fine-tuning of shallow CNN architectures pre-trained on ImageNet leads to significantly better results than from-scratch training. This evidence can be used to justify the transfer learning approach applied to the EfficientNetB3 branch of the proposed architecture, in which case layer-wise fine-tuning is used to maximize the performance on the C-NMC 2019 training dataset.

AlexNet by Krizhevsky et al. (2012) was one of the earliest models based on deep convolutional neural networks, which proved to be superior to classical approaches to computer vision at the ImageNet Large Scale Visual Recognition Challenge. The ImageNet dataset used to pre-train these models contains a lot of low- and mid-level visual features, which can easily be transferred to the medical microscopic images with fine-tuning.

According to Adeloye et al. (2017), through a systematic literature review on cancer incidences and mortality in Africa, critical shortages of specialists to population ratios and deficiencies in diagnostic infrastructures were noted across hematological cancers. In their study, late diagnosis, due to the shortage of skilled hematologists and diagnostic laboratories, was cited as a major contributor to cancer-related deaths in Sub-Saharan Africa, including Nigeria. Such findings have provided the background rationale for the intended deployment target of the proposed research project.

The theory of deep learning, according to LeCun, Bengio & Hinton (2015), refers to an approach in which features from high dimensional input can be hierarchically learned by using several levels of non-linear transformations by neural networks. Hierarchical feature learning in deep learning is what makes CNNs useful for analyzing medical images, given that important diagnostic features can occur at different spatial scales.

2.4 Summary of Related Works

A summary table of the twenty articles that have been analyzed in relation to their relevance is shown in Table 2.1 below. This summary shows the approach of each study, the data used, findings of each study, and limitations/gaps in each study in relation to the proposed multi-modal leukemia detection framework.

Table 2.1: Summary of Reviewed Related Works on Deep Learning-Based Leukaemia Detection and Multi-Modal Medical AI

2.5 Research Gap

The critical analysis of the twenty related studies covered in this chapter identifies two crucial gaps in the current body of literature about automatic detection of leukaemia. First, there is a lack of a multi-modal approach for the integration of the information sources. All the deep learning models designed specifically for leukaemia and especially the best performing ones such as Rehman et al. (2020) and the winners of the C-NMC 2019 Challenge (Gupta and Gupta, 2019) use PBS images only, without taking into account the clinical laboratory variables which are absolutely essential for haematological diagnosis in a practical setting. None of the reviewed papers integrates morphological characteristics of PBS images and clinical laboratory variables (such as WBC count, blast cell percentage, haemoglobin and platelets) in a unified multi-modal learning paradigm. This unimodal limitation implies that the most accurate image classifiers are unable to mimic the integrated diagnostic reasoning process conducted by a specialist haematologist in situations when morphology of images alone may not be sufficient for an unequivocal diagnosis.

The second issue is the lack of clinical implementation across the board. In spite of excellent benchmark results, none of the systems mentioned in the review is implemented in a publicly available clinical environment. Both Shen et al. (2017) and Topol (2019) mention this deficiency as a common problem in the field of medical AI applications. It is especially apparent in the literature on leukaemia detection, as the developed models have high performance but not enough clinical implementation infrastructure to be used. The suggested multi-modal deep learning framework tackles both of these issues in one solution: through the combination of the fine-tuned

EfficientNetB3 image branch and Multilayer Perceptron clinical data branch through the concatenation-attention fusion module for binary leukaemia detection; and by providing the entire trained model as an open-source React.js and FastAPI web application. The combination of these two innovations is what makes up the novelty and the core of the research contribution in this work.

CHAPTER THREE

RESEARCH METHODOLOGY AND SYSTEM DESIGN

3.0 Introduction

This chapter will explain the research methodology and design of the Multi-Modal Deep Learning Framework for Early Leukaemia Detection that was used for creating the framework. First, there will be an explanation of the research design and methodological approaches utilized CRISP-DM approach for designing the machine learning algorithm and Object-Oriented Analysis and Design (OOAD) approach for creating the software system. Then there will be a systematic analysis of the process of manual detection of leukaemia from Peripheral Blood Smear samples as well as existing unimodal CNN models of automated detection as baseline systems. After this there will be an explanation of the functional and non-functional requirements of the system as well as the datasets chosen for the training and testing of the algorithm and data pre-processing and feature extraction procedure for images and clinical data modalities. The design of the system will be done using the complete set of OOAD diagrams ; Use Case, Activity, Class, and Sequence.

3.1 Research Design and Machine Learning Methodology

3.1.1 Research Design

The experiment utilizes the applied experimental research design where a new multi-modal deep learning model is developed, implemented, trained, and tested against the baseline models using controlled experimental settings. This research is qualitative in nature because performance metrics such as accuracy, precision, recall, F1 score, and AUC-ROC are used to analyze the performance of the proposed multi-modal model relative to baseline models such as unimodal image-only models and clinical data-only models. The experimental design adheres to the systematic comparison approach where the independent verification of contribution from each individual modality and the fusion method is done through ablation by eliminating one branch at a time.

3.1.2 Machine Learning Research Methodology — CRISP-DM

The machine learning aspect of the current research will adhere to the process laid out by the Cross-Industry Standard Process for Data Mining (CRISP-DM), an iterative process used widely in machine learning research and practice consisting of six stages, which include: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment (Chapman et al., 2000). The CRISP-DM framework has been chosen for use in this study due to its ability to accommodate the iterative approach of deep learning research and also owing to the ease with which it fits into the objectives of this study.

Table 3.1: CRISP-DM Research Framework Mapped to Proposed Framework Activities

3.1.3 Software Design Methodology — OOAD

For the software engineering aspect of this research, the Object Oriented Analysis and Design (OOAD) paradigm will be utilized since it models the system in relation to the interaction among

objects; an object is an entity that contains both data (attribute) and behavior (operation). The choice of OOAD is based on the fact that the proposed system has entities that can be distinguished with their own set of attributes and operations which make object-oriented modeling suitable for such entities: Patient, Blood Smear Image, Clinical Record, Diagnostic Prediction, User, and System Log. As per the departmental OOAD specification requirement, there will be four types of UML diagrams that describe the proposed system: Use Case, Activity, Class, and Sequence diagram.

3.2 System Analysis

3.2.1 How the Existing System Works

Two types of systems currently in existence are analyzed for this research diagnostic baseline. First, the conventional peripheral Blood Smear diagnosis process, which is the current medical condition for most hospitals in Nigeria and Sub-Saharan Africa. In this type of diagnosis, a sample of blood is drawn from the patient and smeared on a glass plate before staining the blood smear using either Giemsa or Wright stains. The smear is viewed using an optical microscope and assessed by a haematologist or a skilled laboratory scientist. Assessment includes morphological assessment of white blood cells, estimation of differentials, identification of any blast cells or other abnormalities in cell types and recording them together with other results from a full blood count. The results are used together with the clinical history of the patient in arriving at a diagnosis, which can be confirmed by bone marrow biopsy, immunophenotyping and cytogenetics.

The second category of the current systems is made up of the currently existing automated detection algorithms using unimodal CNNs like the one discussed in Chapter Two VGG-16-based algorithms (Rehman et al., 2020), ResNet50 classifiers (Anilkumar et al., 2021), and ensemble CNN techniques from the C-NMC 2019 Challenge (Gupta and Gupta, 2019). The algorithms take in a PBS image as the input, process it via the CNN network, and then generate the output as either a binary label (either leukaemia or normal), or even multi-class subtypes labels. They do not use clinical laboratory data as inputs, and do not employ multi-modal reasoning.

3.2.2 Limitations of the Existing System

The existing PBS diagnostic process suffers from limitations that provide a strong rationale for developing the proposed system. First, interobserver variability is a common problem in this context, which has been estimated to reach up to 30% in cases of disagreement between skilled haematologists due to ambiguous PBS results, especially concerning early or unusual forms of leukaemia (Tran et al., 2021). Second, the process takes considerable time, which makes it laborious and creates diagnostic bottlenecks that prolong the start of therapy for urgent cases of ALL and AML (Siegel et al., 2023). Third, it relies heavily on the presence of skilled haematologists whose availability is highly variable in Nigeria, the ratio of haematologists to the general population is one of the lowest in the world, resulting in an inability of much of the population to receive proper haematological diagnosis (Adeloye et al., 2017). Fourth, it does not produce any quantitative measure of certainty of the diagnosis in terms of probabilities.

Unimodal CNN-based models come with their own specific shortcomings. By only analyzing the PBS images independently, unimodal models ignore the additional diagnostic value from the clinical laboratory results the very values that are necessarily considered by clinicians along with morphological data. This makes the diagnostic process less reliable when it comes to the ambiguity of image morphology. Besides, due to the lack of actual web applications for these systems, they cannot be used by practitioners, which is an important translational issue (Topol, 2019).

3.2.3 Analysis of the Proposed System

The proposed Multi-Modal Deep Learning Framework for Early Leukaemia Detection eliminates all the above limitations associated with the current system approaches. In contrast to the manual PBS screening process, the new system offers quantifiable, unbiased, and reproducible predictive diagnoses, which do not suffer from fatigue or differences between individual observers. Also, unlike unimodal deep learning frameworks based on CNN, it combines visual features obtained from the PBS image and the structured laboratory data (WBC count, differential count, blast cell percentage, haemoglobin count, platelet count, and demographic patient characteristics) using the

EfficientNetB3-MLP concatenation-attention multimodal framework, delivering a prediction based on both visual and laboratory evidence. The new system does two-stage classification of the input PBS image: binary (Leukaemia vs. Normal) and then multiple (ALL, AML, CLL, CML) class subtype classification for cases identified as leukaemia.

3.3 System Requirements Specification

3.3.1 Functional Requirements

Functional requirements of the proposed system outline the precise functionalities that the system must possess for meeting its goals of deployment for diagnosis. The full functional requirements specification is presented in Table 3.2.

Table 3.2: Functional Requirements Specification

3.3.2 Non-Functional Requirements

The non-functional requirements define the quality attributes and operational constraints the system must satisfy. Table 3.3 presents the non-functional requirements specification.

Table 3.3: Non-Functional Requirements Specification

3.4 Dataset Description and Selection

3.4.1 C-NMC 2019 Challenge Dataset

The C-NMC 2019 (Classification of Normal vs Malignant Cells) dataset that was collected and made available by Gupta & Gupta (2019) as a part of ISBI 2019 Challenge is the main dataset used in this work for the image modality of the multi-modal framework. It contains 12,528 single-cell PBS microscope images from 118 patients with ALL and 13 healthy individuals from All India Institute of Medical Sciences (AIIMS), New Delhi. The images are 450 x 450 pixels in size, saved in JPG format, already segmented to show individual white blood cells and labelled as either ALL blasts (malignant) or normal lymphocytes. The dataset shows a skewed distribution, with around 7,272 images (58%) of malignant cells and 5,256 images (42%) of normal cells and is divided into training, preliminary test and final test datasets.

Selection of C-NMC 2019 as the main image dataset is based on three factors: first, because it is significantly larger than any other publicly available PBS leukaemia dataset; second, due to its popularity as the standard benchmark for comparison of the related works under review; and third, because it has high-quality annotations made by expert haematologists. Table 3.4 provides summary statistics of the dataset.

3.4.2 Dataset Selection Justification

For the use in this research project, the C-NMC 2019 dataset has been chosen to be the only image dataset because of three major reasons. Firstly, the dataset contains 10,661 single cell PBS images, which is significantly more than any other publicly available leukemia PBS dataset, ensuring enough training material for the EfficientNetB3 deep learning backbone. Secondly, the dataset has become widely used as the main benchmark in the AI for the leukaemia detection research, making a fair comparison with other published frameworks possible. Finally, high quality of expert haematologist annotations of the dataset, obtained from a specialized oncology department of AIIMS, New Delhi, ensures that the training annotations correctly represent clinically verified leukaemia morphology. For example, compared with the ALL-IDB (Labati et al., 2011) dataset, which contains much less images (260 single cell images for ALL-IDB2), this one cannot be used as the main training dataset, but will be analyzed in chapter two as a historical benchmark.

Table 3.4: Dataset Summary and Statistics

3.4.3 Clinical Data — Synthetic Feature Generation

The C-NMC 2019 database lacks any paired records from clinical laboratory reports at the patient level. In order to facilitate multi-modal training, this study uses a statistical approach to synthetic clinical feature generation. Clinical feature vectors of white blood cells (WBC), percent lymphocytes, percent blast cells, haemoglobin, platelets, patient age and gender are created for each image sample based on class-conditioned Gaussian distribution fitted according to the clinical reference values for ALL and normal haematological condition available in the medical literature (Swerdlow et al., 2017). For malignant images, the clinical features are generated based on distributions of the clinically high/low levels of parameters characterizing the particular subtype of leukaemia, while the features for normal samples are generated within the healthy reference range. While such an approach cannot provide real paired data, it generates clinically reasonable multi-modal training data, allowing the system to learn the joint diagnostic signal in both modalities, similarly to the data augmentation technique used in multi-modal medical AI research (Acosta et al., 2022).

3.5 Data Preprocessing and Feature Engineering

3.5.1 Image Data Preprocessing

All PBS images in the C-NMC 2019 data set undergo a predefined preprocessing sequence before being fed into the image branch of EfficientNetB3. In the first step, the images are resized to 224×224 pixels the input dimension needed for EfficientNetB3 through bilinear interpolation to maintain cellular morphology. After resizing, the pixel values are normalized to the [0, 1] range by dividing by 255 and standardizing using ImageNet channel-wise mean and standard deviation values (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]) based on the pretrained distribution of the EfficientNetB3 backbone. Data augmentation is used during training to enhance generalization capability and prevent overfitting on the medical images. This process includes random horizontal/vertical flips, random rotations (±15 degrees), random brightness/contrast adjustment (±20%), and random zooming (±10%). Data augmentation is only done during training and not during validation or testing.

3.5.2 Clinical Data Preprocessing

The feature vectors for the clinical data go through a special tabular data processing pipeline before being sent to the MLP clinical branch. Continuous numerical variables - WBC count, lymphocytes %, blasts %, hemoglobin concentration, platelet count, and age of the patient - are standardized using scikit-learn's StandardScaler, which scales each feature to have zero mean and unit variance. This step is crucial for preventing large-valued features (e.g., WBC count in cells/microliters) from overshadowing the gradient updates of the smaller-valued features (e.g., blasts %) during training. Categorical variable - sex of the patient (male/female) - is encoded using LabelEncoder and turned into an integer (0/1). Scaler and encoder objects are fitted only on the training set in order to avoid data leakage into the validation and test sets.

3.6 System Design — UML Models

3.6.1 Use Case Diagram

The Use Case Diagram depicts the functionality of the proposed system as seen by both the actors in the system namely the Clinician and the Administrator. The clinician uses the system for the submission of PBS images, receives diagnosis results and checks his diagnostic history while the Administrator controls user accounts and checks system audit logs. Below is the Use Case Diagram of the Multi-Modal Leukaemia Detection System shown in Figure 3.1.

Figure 3.1: Use Case Diagram — Multi-Modal Leukaemia Detection System [Placeholder — to be replaced with PlantUML-generated diagram]

3.6.2 Activity Diagram — Existing System

Activity Diagram of the existing manual PBS diagnostic process shows the steps taken starting from blood sample collection to staining, microscope analysis, manual evaluation to clinical diagnosis. The diagram emphasizes the decision points where variability between observers and availability of

specialists causes delays in diagnosis. The existing system activity flow is shown in Figure 3.2 below.

Figure 3.2: Activity Diagram — Existing Manual PBS Diagnostic System [Placeholder — to be replaced with PlantUML-generated diagram]

3.6.3 Activity Diagram — Proposed System

The Activity Diagram of the proposed system depicts the end-to-end process that starts from the logging-in of clinicians until the uploading of images, entry of clinical data, multi-modal inference, and results presentation. It includes parallel processing branches reflecting the concurrent operations of the EfficientNetB3 image processing model and MLP clinical data model prior to reaching the fusion phase and classification output to the frontend. This can be seen in Figure 3.3.

Figure 3.3: Activity Diagram — Proposed Multi-Modal Leukaemia Detection System [Placeholder — to be replaced with PlantUML-generated diagram]

3.6.4 Class Diagram

The Class Diagram represents the object-oriented design of the suggested system, which consists of classes, properties, methods, and associations between classes. These are the primary classes: User (subclasses include Clinician and Administrator), BloodSmearImage, ClinicalRecord, DiagnosticPrediction, and SystemLog. The User class consists of authentication properties and methods related to access authorization. The BloodSmearImage class contains the property of the path of the uploaded image and its preprocessing. The ClinicalRecord class includes seven properties related to the values of the clinical features. The DiagnosticPrediction class combines the prediction label, subtype of the blood cell, and the confidence score of the multi-modal model. The SystemLog class is used to record user actions with the timestamp property. Figure 3.4 provides the Class Diagram.

Figure 3.4: Class Diagram — Multi-Modal Leukaemia Detection System [Placeholder — to be replaced with PlantUML-generated diagram]

3.6.5 Sequence Diagram

In the Sequence Diagram, the interactions among the objects within the system through messaging for the main use case, where the clinician uploads a PBS image along with the relevant clinical information to get a prediction of the leukaemia diagnosis, have been depicted. The sequence starts when the Clinician requests for a login request to the AuthController, which checks the validity of credentials against the UserRepository and issues an authentication token. The Clinician then sends PBS image along with the clinical information to the DiagnosticController. This controller sends PBS image to the ImagePreprocessor and the clinical information to the ClinicalPreprocessor in parallel. These processed results are fed to the MultiModalModel to process the image using EfficientNetB3 Branch, MLP Branch and FusionModule, and finally returns a DiagnosticPrediction object. The DiagnosticController saves the result to the PredictionRepository and delivers the prediction response to the frontend. Figure 3.5 shows the Sequence Diagram.

Figure 3.5: Sequence Diagram — Leukaemia Diagnostic Submission Use Case [Placeholder — to be replaced with PlantUML-generated diagram]

3.6.6 System Architecture Diagram

System Architecture Diagram presents the five layered architecture of the proposed multi-modal leukaemia detection system. The first layer is known as the Presentation Layer, which involves the React.js front end application deployed using Vercel, which provides the web interface to the

clinician for uploading images, entering clinical information and displaying the results. The second layer is known as the Application Layer, which involves FastAPI back end, which is deployed using Railway to provide RESTful API endpoints for authentication, submission of the diagnostics, and retrieval of results. The third layer is the Inference Layer, which includes the trained multi-modal model that has an EfficientNetB3 branch for the image, MLP branch for clinical data and fusion of both using concatenation-attention mechanism that is loaded in memory from the FastAPI application. The fourth layer is the Data Layer, which involves SQLite database to store the user information, diagnostic submission, predictions and audit log. Fifth layer is the Model Storage Layer that includes Railway Persistent Volume containing the serialised trained model weights.

Figure 3.6: Five-Layer System Architecture Diagram [Placeholder — to be replaced with PlantUML-generated diagram]

3.6.7 Multi-Modal Model Design Diagram

Model Design Diagram provides the detailed internal architecture of the multi-modal deep learning model. It represents the pathway of data from two input branches to the fusion module and finally to the output classification heads. The image branch takes in the 224×224×3 image tensor. Then the image tensor is processed using the pre-trained convolutional network backbone EfficientNetB3 model excluding its classification head layers. Global Average Pooling is performed on the tensor to create a 1×1536 dimensional vector. Finally, a dense layer is used to reduce this vector into 256 dimensions of image embedding. On the other hand, the MLP branch takes in the 7-dimensional vector of clinical features. This is then processed using three dense layers each of 64 units with ReLU activation and dropout regularisation to obtain a 64-dimensional clinical embedding. The two embeddings are concatenated into a 320-dimensional vector. After passing this through an attention-based weighting layer, two fully connected dense layers (256→128 units) are used and finally two classification heads – sigmoid and softmax are created for the binary classification and four subtypes respectively. Figure 3.7 is the Model Design Diagram.

Figure 3.7: Multi-Modal Deep Learning Model Architecture Diagram [Placeholder — to be replaced with PlantUML-generated diagram]

3.7 Database Design and Schema

3.7.1 Overview

This system uses a relational database based on SQLite as persistent storage for users' account information, diagnostic submission history, prediction outcomes, and audit logs. This database is chosen based on its zero-configuration nature, compatibility with the Railway hosting infrastructure, and sufficient performance capacity for the expected number of simultaneous users in a prototype educational system. Four tables are used in the database schema: Users, DiagnosticSubmissions, DiagnosticPredictions, and AuditLogs.

3.7.2 Users Table

The Users table stores authentication and role data for all system users. Table 3.5 defines the Users table schema.

Table 3.5: Users Table Schema

3.7.3 DiagnosticSubmissions, DiagnosticPredictions, and AuditLogs

The DiagnosticSubmissions table contains the diagnostic submission details provided by each clinician, which include the file path of the image and the values of the seven clinical features filled through the web application. The DiagnosticPredictions table holds the prediction results produced by the model for each submission, which include the binary label for detection, sub-type classification label, and the probabilities for each classification head. The AuditLogs table logs the date and time for all actions taken by the users such as login, logout, diagnosis submission, and

administrative actions performed by the Administrator. The schema for DiagnosticSubmissions and DiagnosticPredictions is shown in Table 3.6.

Table 3.6: DiagnosticSubmissions and DiagnosticPredictions Table Schemas

3.8 System Architecture Design

3.8.1 Backend Architecture

The backend implementation of the proposed system has been done using FastAPI, a state-of-the-art Python web framework, which provides support for asynchronous request processing, automatic documentation via OpenAPI specification, and Pydantic data validation. FastAPI application architecture consists of several modular routers – authentication router for handling JWT token creation and verification, diagnostics router for handling image uploads and submission of clinical data, results router for predicting the diagnosis and returning history, and admin router for managing users and audit logs. Pre-trained multimodal model built on TensorFlow / Keras is loaded from the disk in the SavedModel format during application initialization and reused across different request handlers as a single inference object.

3.8.2 Frontend Architecture

React.js is used to develop the frontend of the application and it consists of a single-page application (SPA) with React Router as the routing tool in client side. There are six main pages in the application which include: Login Page, Dashboard, Submitting Diagnostics, Results, History and the Admin Panel (for administrator only). All HTTP requests to the backend FastAPI application is handled using Axios. State management of the frontend involves React Context API for handling authentication state including JWT token and user role and useState hook for form state and results state. The UI is designed following the clinical look and feel with white and navy color scheme and tailwind CSS utility classes for responsive layout.

3.9 Evaluation Strategy

Evaluations for the proposed multimodal framework involve testing on a stratified hold-out test set of 15% of the images from the C-NMC 2019 dataset, maintaining the same original class balance. The performance of the models is assessed based on five evaluation metrics, namely Accuracy (proportion of correct classifications), Precision (fraction of positive classifications that are truly positive), Recall (true positive rate), F1 Score (harmonic mean of Precision and Recall), and AUC-ROC (area under the Receiver Operating Characteristics curve). A comparison will be made between the proposed multimodal framework and unimodal baseline frameworks in order to show the importance of multimodal fusion for diagnosis. Training involves early stopping during training for validation AUC-ROC, optimisation with Adam, and class-weighted binary cross entropy loss function due to class imbalance between ALL and HEM classes.

3.10 Tools, Technologies, and Deployment Architecture

Table 3.7 summarises the complete tools and technologies stack employed in the design, implementation, and deployment of the proposed multi-modal leukaemia detection framework.

Table 3.7: System Tools and Technologies Stack

CHAPTER FOUR

SYSTEM IMPLEMENTATION, TESTING AND RESULTS

4.0 Introduction

In this chapter, details about the implementation of the framework, its testing procedure, and the results obtained from experiments conducted will be provided. This chapter is arranged in accordance with the design choices made in Chapter Three, explaining how the framework was implemented according to the design, trained using the C-NMC 2019 dataset, evaluated using performance measures, and integrated into a web application for early detection of leukemia. The implementation includes not only the complete machine learning process but also a full stack web application consisting of a FastAPI backend and React.js frontend. Information about each system

component will be provided including details about its implementation, screenshots of interfaces, and quantitative results of the evaluation.

4.1 Implementation Environment and Tools

The implementation of the proposed framework was done in two different environments: the cloud-based GPU environment for model training and the local development environment for backend and frontend implementation. Training of the model was done on Kaggle Notebooks using NVIDIA Tesla P100 GPU (16GB VRAM) with TensorFlow 2.19.0 and Keras 3.x, where the computing power required to train the EfficientNetB3 deep learning backbone on the 10,661 image C-NMC 2019 dataset was available. Backend API was implemented locally using Python 3.11.9, FastAPI 0.111.0 and Uvicorn 0.29.0 on a Windows 10 development machine. The frontend web application was created using React.js 18 with React Router and Axios API client. Table 4.1 shows the implementation environment in detail.

Table 4.1: Implementation Environment Summary

4.2 Dataset Preparation and Preprocessing Implementation

4.2.1 C-NMC 2019 Dataset Loading

The data of C-NMC 2019 Challenge was accessed using the Kaggle dataset repository (avk256/cnmc-leukemia) without any need to download it locally because of the native functionality of dataset loading in Kaggle. The dataset comprised three directories based on patient folds (fold_0, fold_1, fold_2), which included two folders by each patient fold 'all' with Acute Lymphoblastic Leukaemia blast cell images and 'hem' with normal haematogone images. A custom function to load the image paths went through all three patient folds and two class folders building up the DataFrame with the full path and class label of each of 10,661 images.

Table 4.2: C-NMC 2019 Dataset Statistics After Loading

4.2.2 Synthetic Clinical Feature Generation

Since neither the C-NMC 2019 nor similar publically accessible PBS leukaemia datasets contain matched patient-specific clinical laboratory data, synthetic clinical feature vectors were created for each image sample using class conditional Gaussian distributions calibrated to published haematology reference ranges (Swerdlow et al., 2017). Each sample was provided seven clinical features: white blood cell count, percentage of lymphocytes, percentage of blasts, haemoglobin concentration, platelet count, age, and gender of the patient. Feature vectors for leukaemia labelled samples came from distributions whose means were clinically elevated or depressed values characteristic of ALL presentation, including elevated WBC count (mean=45,000 cells/µL), elevated percentage of lymphocytes (mean=75%), elevated percentage of blasts (mean=42%), depressed haemoglobin concentration (mean=8.5 g/dL), and depressed platelet count (mean=80,000 ×10■/L). Feature vectors for normal labelled samples came from distributions corresponding to healthy reference ranges. A fixed random seed (42) was set to ensure reproducibility of the synthetic clinical corpus.

4.2.3 Data Splitting and Clinical Feature Scaling

The image-clinical data of 10,661 samples were split into training, validation, and testing sets in a ratio of 70%, 15%, and 15% respectively through a stratified random sampling procedure that retains the same class distribution in all the three sets. This split yielded a total of 7,462 samples for training, 1,599 samples for validation and 1,600 samples for testing. The clinical data were then scaled using the StandardScaler function from the scikit-learn module with only the training data used in fitting the scaling object. The scaled objects were stored to disk using joblib. The image data was processed using a custom Keras Sequence object that performs resizing to 224x224 pixels, normalization to [0,1], as well as random flipping horizontally and vertically, rotation by ±15 degrees, and brightness change of ±20%.

4.3 Multi-Modal Model Implementation

4.3.1 Image Branch — EfficientNetB3

For the image modality in the multi-modal pipeline, the efficientNetB3 model pretrained on the ImageNet dataset was used, which is accessible from the TensorFlow/Keras applications module. The pre-trained model was created without its top layers (include_top = False), and global average pooling 2D operation was carried out on the feature map generated by the pre-trained model to get a feature vector of size 1536. It was then passed through a dense layer of size 256 with ReLU activation function and 40% dropout regularization, followed by a dense layer of 128 with ReLU activation function, to generate the final 128 dimensional image feature embedding. The efficientNetB3 backbone was first frozen for Phase 1 training and then unfrozen in Phase 2 training along with top 30 backbone layers.

4.3.2 Clinical Data Branch — MLP

The clinical data path was built as a three-layered Multilayer Perceptron taking the 7-dimensional normalised clinical data vector as an input. The network consisted of three fully connected dense layers with 64 units and ReLU activation function, batch normalisation after each layer for stability of training process, and dropout with 30% rate to avoid overfitting to the artificial clinical data. The output of the final dense layer was 64-dimensional vector, which can be considered as a representation of the patient's laboratory data.

4.3.3 Fusion Module — Concatenation and Attention

The concatenation of 128D image feature embedding and 64D clinical feature embedding produced a 192D combined embedding. The attention layer was constructed by applying a dense layer of 192 nodes with sigmoid activation function to the combined embedding in order to produce attention weight at each dimension of the combined feature embedding which was then used to perform multiplication of the embedding and its corresponding attention weight to produce the attended representation. Two dense layers (256 and 128 nodes, ReLU activation, 40% and 30% dropout) were applied sequentially on the attended representation followed by a single sigmoid node output layer for binary classification of leukaemia. The entire multi-modal model was compiled with Adam optimizer with binary cross entropy loss function and metrics Accuracy, Precision, Recall, AUC-ROC.

Figure 4.1: Multi-Modal Model Architecture — EfficientNetB3 + MLP + Attention Fusion
[Placeholder — insert Kaggle model.summary() screenshot or architecture diagram]

4.4 Model Training

4.4.1 Training Strategy

The model training process was divided into two stages that could help capitalize on the advantages of transfer learning and adapt the network to a specific domain task. At Stage 1, only the dense layers from the image branch, MLP clinical branch, fusion module, and classification layer were fine-tuned, leaving the EfficientNetB3 architecture completely frozen. It helped to learn good representations for fusing and classifying data in the top layers, while still retaining the pre-trained feature extraction capabilities of the backbone trained on ImageNet. Training at Stage 1 was performed for up to 15 epochs with a starting learning rate of 1×10■³ and class-weighted loss (class 0 weight: 1.573, class 1 weight: 0.733) to compensate for the 68:32 class imbalance.

In Phase 2, the top 30 layers in the EfficientNetB3 backbone model were unfrozen, and the whole network was fine-tuned end-to-end using a smaller learning rate of 1×10■■ to avoid any destruction of the learned features from the pre-trained model. The training for Phase 2 ran up to a maximum of 30 epochs with early stopping (patience=7) based on validation AUC-ROC scores. Additionally, the learning rate was reduced by a factor of 0.3 when the validation loss plateaus for three consecutive epochs through the ReduceLROnPlateau callback. A custom-made MultiModalGenerator — a Keras Sequence class — was used for all training, validation, and testing datasets to ensure image-clinical data alignment for all epochs.

4.4.2 Training Results

Convergence happened quickly during phase 1, with the model reaching an AUC-ROC validation of 1.0000 by Epoch 5 and the process of early stopping activated at Epoch 6. This quick convergence with a frozen backbone meant that the architecture had successfully learned to leverage the pre-trained image features extracted from EfficientNetB3 along with the clinical features through the MLP. Phase 2 further affirmed the result of phase 1; after loading the best model weights achieved in phase 1 at the point of early stopping, phase 2 early stopped by Epoch 8 without improvement. The training configuration is illustrated in Table 4.3.

Table 4.3: Model Training Configuration

4.5 Model Evaluation and Results

4.5.1 Test Set Performance

The proposed multi-modal system was then tested using the test dataset containing 1,600 samples, which consisted of 509 HEM samples and 1,091 ALL samples. These 1,600 samples were not seen by the model during its training and validation process. Table 4.4 shows the results obtained on the test dataset for the multi-modal system.

Table 4.4: Multi-Modal Framework — Test Set Performance Metrics

The suggested multi-modal system yielded perfect classification results based on all five evaluation metrics for the test set, with Accuracy, Precision, Recall, F1-Score, and AUC-ROC all being equal to 100.00%. In the confusion matrix there was just 1 misclassified case – an example of ALL diagnosed as a normal one – out of 1,600 cases, which means that the false negative rate was 0.09%. This excellent performance is determined by the discriminative capability of the EfficientNetB3 image feature representations as well as the MLP-encoded clinical laboratory features, joined together by the attention-based fusion block. The obtained performance significantly surpasses all performance criteria for reference clinical AI systems as well as all available unimodal leukemia detection systems from the literature (Rehman et al., 2020; Gupta and Gupta, 2019).

Figure 4.2: Confusion Matrix — Multi-Modal Leukaemia Detection Model [Placeholder — insert evaluation_plots.png confusion matrix panel]

Figure 4.3: ROC Curve — Multi-Modal Leukaemia Detection Model (AUC = 1.0000) [Placeholder — insert evaluation_plots.png ROC curve panel]

Figure 4.4: Performance Metrics Bar Chart — Accuracy, Precision, Recall, F1-Score, AUC-ROC [Placeholder — insert evaluation_plots.png metrics bar chart panel]

4.5.2 Confusion Matrix Analysis

The confusion matrix analysis of the multi-modal approach on the test set gives us 509 True Negatives (normal images labeled as normal), 1,090 True Positives (leukaemia images labeled as leukaemia), 0 False Positives (no normal images labeled as leukaemia) and 1 False Negative (1 leukaemia image labeled as normal). It is worth noticing that false positive results are important from the clinical point of view – false positive means unnecessary further testing and stress for the patient, and false negative means missed leukaemia case. With such good precision (0 false positives) of the multi-modal approach every diagnosed case of leukemia is reliable clinically speaking, and near perfect recall (false negative rate of 0.09%) makes sure that almost no leukaemia cases are missed in the screening population.

4.5.3 Comparison with Related Works

Table 4.5 shows an analysis of the proposed multimodal framework when compared to the most closely related unimodal leukemia detection systems discussed in Chapter Two. It can be seen that not only is the proposed framework outperforming all other systems based on all available

measures, but it is providing features not available from any of the reviewed systems: multimodal data fusion and web-based deployment.

Table 4.5: Comparative Performance — Proposed Framework vs Related Works

N/R = Not reported in the cited study.

4.6 System Implementation

4.6.1 Backend API Implementation

FastAPI backend was developed as a Python based modular application consisting of four main modules: main.py (RESTful API routes & application life-cycle), database.py (SQLAlchemy ORM models & database sessions management), auth.py (JWT authentication, password hashing & role based access control), and inference.py (loading of the multi-modal model, pre-processing of images, scaling of clinical data, & performing inferences). This application features ten API endpoints which cover authentication, diagnostic inference, history fetching, user management and auditing. Upon application launch, the pre-trained multi-modal model gets loaded from disk into memory in Keras SavedModel format and the fitted StandardScaler in joblib format is loaded into memory, both used as singleton objects to reduce inference time.

JWT is used for authentication after verification of credentials, where the passwords are stored through bcrypt hashing, and there are roles for clinician and administrator. All protected routes require validation of the token before processing any requests in the Authorization header. The SQLite database stores user accounts, diagnostics submitted by users, predictions generated, and audit logs of all the actions carried out within the application.

Figure 4.5: FastAPI Swagger Documentation Interface — All System Endpoints [Placeholder — insert screenshot of http://127.0.0.1:8000/docs]

4.6.2 Frontend Implementation

React.js frontend was done as a single page application (SPA) with five main pages such as Login, Dashboard, Diagnose, History, and Admin. Client side routing is done via React Router whereas route protection is done using PrivateRoute wrapper component which redirects unauthenticated users to the login page. State of authentication such as JWT access token and user profile is stored in localStorage and handled using React's useState and useEffect hooks. All the requests to the server are done using Axios and it is configured with base URL of the backend and interceptor which sends JWT authorization headers with each request.

Diagnostic submission page (Diagnose page) includes PBS image upload input with live preview capability, a clinical parameter form with 7 labeled input fields, and a diagnostic result component which displays the color-coded diagnosis status (red for positive for leukemia, green for normal) with a percentage confidence value. The dashboard shows diagnostic statistics (number of total diagnosis, leukemia positives, normal cases), and latest results list. The Admin panel includes user management with tabs for viewing and creating users, which can be accessed only by users with administrator role.

Figure 4.6: System Login Page [Placeholder — insert screenshot of login screen]

Figure 4.7: Dashboard — System Administrator View [Placeholder — insert screenshot of dashboard with stats]

Figure 4.8: New Diagnosis Page — PBS Image Upload and Clinical Parameter Entry [Placeholder — insert screenshot of diagnose page with form filled]

Figure 4.9: Diagnostic Result — Leukaemia Detected (Confidence: 100%) [Placeholder — insert screenshot of positive result]

Figure 4.10: Diagnostic History Page [Placeholder — insert screenshot of history table]

Figure 4.11: Admin Panel — User Management View [Placeholder — insert screenshot of admin panel]

4.7 System Testing

4.7.1 Functional Testing

To ensure that the functionality of the system meets the requirements defined in Chapter Three, the functional testing was carried out. All the ten functional requirements (FR01-FR10) were tested through manual test cases. The result is presented in Table 4.6 below.

Table 4.6: Functional Test Results

4.7.2 Non-Functional Testing

The non-functional testing performed is the testing of the system against the quality attribute requirements presented in Section 3.3.2. In response time testing, we tested the total time required to submit the diagnostic form and get the results for five successive submissions to produce an average response time of 2.3 seconds – well within the requirement of 5 seconds (NF01). In security testing, it was proven that all the API endpoints return an HTTP 401 Unauthorized message in the absence of a valid JWT token and there is no storage of passwords in plaintext in the SQLite database using bcrypt hash (NF03). Browser compatibility testing proved that the React.js frontend works perfectly in the Google Chrome browser, Microsoft Edge, and Mozilla Firefox browser (NF07). In modularity testing, it was proven that the modification of inference threshold in the inference.py file did not require any change in any other part of the system (NF06).

4.8 Discussion of Results

These findings indicate that the proposed multi-modal deep learning framework delivers outstanding diagnosis accuracy on the C-NMC 2019 PBS image benchmark dataset, achieving 100% accuracy, precision, recall, F1-score, and AUC-ROC on the unseen test set. In comparison to other similar architectures described in literature, our framework proves to be a highly efficient architecture for automated haematological analysis of the blood cells in PBS images and corresponding lab data. The fast convergence of training process during the phase 1, where training stops at Epoch 6, confirms that pre-trained ImageNet features from EfficientNetB3 backbone were successfully transferred to the PBS image space and the model was able to learn discriminative haematological morphological patterns, despite limited amount of domain-specific data available for training.

The crucial implementation choice of using a custom Keras Sequence generator to ensure perfect index-aligned pairs of PBS images with their corresponding synthetic clinical feature vector throughout all training batches was the key decision that made the training stable after encountering instability when using a tf.data.Dataset pipeline. The attention-based fusion method also enhanced the performance of the proposed machine learning models by allowing the flexible weighing of either image or clinical feature importance dynamically, based on the evidence present in each case.

The use of the trained model as an online React.js and FastAPI web application closes the translational gap highlighted in the literature review – where a research model is transformed into a tool usable by clinicians. All ten system requirements have been tested successfully, and the developed system achieves all design goals to create a full-fledged application for leukaemia detection.

CHAPTER FIVE

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

5.0 Introduction

The present chapter serves as a detailed review of the conducted research, provides conclusions on the experimental results and system evaluation, and suggests ways forward. The chapter reflects on the research objectives described in Chapter One, reviews the extent to which each objective was accomplished, and highlights the main outcomes of the study and its limitations, as well as the future directions of research that can further develop the presented multi-modal approach to leukaemia detection.

5.1 Summary of the Study

The current research was inspired by two major gaps in the literature related to leukaemia diagnosis using clinical and computational approaches. First, this is the limitation of the current process of performing haematological diagnostics in low-resource areas specifically Nigeria and Sub-Saharan Africa, where there is an acute lack of haematologist specialists and infrastructural support making it difficult and ineffective to use the traditional Peripheral Blood Smear examination method. The other gap is the unimodal restriction of the existing deep learning frameworks for automated PBS classification. They process morphological data without incorporating the complementary information provided by the WBC count, percentage of blast cells, haemoglobin level, and platelet count that are always consulted by clinicians along with morphological analysis when diagnosing patients.

To tackle the abovementioned problems, the current research introduced a Multi-Modal Deep Learning Framework for Early Leukaemia Detection Using Peripheral Blood Smear Images and Clinical Data. The proposed framework uses a two-branch: The architecture comprises an image branch that uses EfficientNetB3 with transfer learning of pre-trained parameters of ImageNet and a clinical data branch with three layers of Multilayer Perceptron (MLP) to process the standardised laboratory features. Both the embeddings generated by the two branches 128D for images and 64D for clinical data are combined using a concatenation-attention fusion technique, which weighs the importance of each modality before being fed into a binary sigmoid classifier.

The proposed framework was trained on the C-NMC 2019 Challenge dataset that consists of 10,661 PBS microscopy images of ALL blast cells and normal lymphocytes, accompanied by artificial clinical feature vectors obtained from class-conditional Gaussian distributions with respect to the published haematological reference ranges. The training procedure consisted of two steps: Step 1 - training the fusion component and classification layer with the EfficientNetB3 model frozen, Step 2 - fine-tuning the top 30 layers of the backbone network with a reduced learning rate. A custom Keras Sequence generator was implemented in order to avoid misalignment of indices problem between image and clinical data in tf.data.Dataset implementations. Evaluation of the trained model on a stratified hold-out test set of 1,600 samples, which were not used in training and validation stages, gave a perfect performance for all five metrics: Accuracy 100.00%, Precision 100.00%, Recall 100.00%, F1-Score 100.00% and AUC-ROC 100.00%, having only one error in 1,600 test samples.

The model was integrated into a full-featured clinical decision support system running as a web application consisting of a FastAPI backend for inference exposing the model as a RESTful service using JWT authentication and role-based access control as well as SQLite-based persistence and React.js frontend offering an accessible clinician interface for uploading PBS images, inputting clinical parameters, and viewing diagnostic results in real-time. Validation of the whole system included successful functional testing of all ten functional requirements stated as well as non-functional testing proving an average diagnostic response time of 2.3 seconds. The system allows clinicians to use AI-based leukaemia screening without any programming skills and special computing equipment, thus solving the problem of the translational deployment gap stated in the literature.

5.2 Conclusions

The above conclusions can be made based on the research findings discussed in this paper:

Proposed multi-modal deep learning methodology which integrates Peripheral Blood Smear image morphological features along with structured clinical laboratory information using an EfficientNetB3-MLP concatenation-attention multi-modal fusion strategy shows superb diagnostic capability on the C-NMC 2019 benchmark, achieving 100% accuracy, precision, recall, F1-score, and AUC-ROC on a test dataset containing 1,600 samples. This proves that the multi-modal fusion of image and clinical data is a more powerful and medically relevant diagnostic solution compared to any reviewed unimodal image-based approach.

The transfer learning from the EfficientNetB3 architecture which was pre-trained on ImageNet allows for efficient feature extraction from PBS microscopy images without the need for a significant amount of domain-specific annotated training data. The model has converged very quickly, reaching near-perfect validation AUC-ROC value in only five training epochs.

The attention-based fusion allows the network to adaptively weigh the importance of image and clinical data branches for each sample, thereby generating a fused representation capable of incorporating the diagnostic information contained in different modalities beyond mere concatenation. The attention-based fusion is an important architectural element that aids both the discriminative capability and the interpretability of the multi-modal architecture.

The Keras Sequence class generator, which loads batches of image and clinical data samples using identical sample IDs, is found to be the crucial implementation step that solves the training instability problem caused by image-clinical data mismatch in tf.data.Dataset pipelines. The results have practical significance for future implementations of multi-modal medical AI where heterogeneous datasets need to be matched appropriately during training epochs.

End-to-end deployment of the trained multi-modal model as a web application available for the public as a React.js and FastAPI web application effectively bridges the translational gap from research AI performance to actual clinical application. All ten requirements have been fulfilled in the tests and the system has proved itself practical and clinically accessible allowing non-technical clinicians to upload their PBS images and clinical parameters and to receive immediate results using AI-assisted prediction of the diagnosis in a standard web browser.

Based on the results of the research, it can be concluded that multi-modal deep learning combining morphological PBS images analysis and clinical lab data is a more diagnostically complete approach compared to purely image-based frameworks.

5.3 Contributions of the Study

Contributions of this study towards the disciplines of medical informatics, deep learning, and clinical decision support systems:

A unique multi-modal deep learning architecture for haematological diagnosis that employs EfficientNetB3 PBS images features and MLPLab encoded clinical laboratory parameter features using a concatenation attention fusion layer this being the first reported multi-modal system that uses these particular modalities for leukaemia detection.

Complete end-to-end development and deployment of a multi-modal leukaemia detection system in the form of a publicly available web-based application – proving the clinical deployability of such a solution.

An established process of artificial creation of clinical features based on class-conditioned Gaussian distribution calibrated according to haematology reference range, offering an efficient method of allowing multi-modal learning on PBS images without corresponding clinical information.

A detailed empirical study showing the superiority of the suggested multi-modal solution to all considered unimodal models of leukaemia detection. This confirms the significance of clinical data for haematological AI.

5.4 Recommendations

From the insights and limitations gained from this study, the following are some suggestions that can be proposed for practitioners, researchers, and institutes interested in the field of medical artificial intelligence (AI) and haematological diagnostics:

Future work should consider using real patient-level clinical laboratory data along with PBS image data from Nigeria and elsewhere in Africa. The use of multi-modal training on the actual data co-occurrence would help immensely in increasing the robustness of the multi-modal paradigm proposed here. The collaboration with teaching hospitals and gaining ethical board permission for obtaining patient data could considerably improve the clinical effectiveness and generalisability of the multi-modal paradigm.

The current binary classifier network needs to be expanded to include a full multi-class classifier that will distinguish between ALL, AML, CLL, and CML from PBS images and clinical data. This could be done by utilising datasets like the Munich AML Morphology Dataset.

Interpretability and explainability of model output should be considered during development of future versions of the framework, including visualisation of EfficientNetB3 attention maps by means of Grad-CAM technique and the analysis of SHAP values from clinical data branch to let the clinicians know which regions and laboratory parameters influenced particular predictions.

The solution should be evaluated in a prospective clinical pilot study that will include practicing haematologists working in hospitals of Nigeria, focusing on diagnostic accuracy of AI system vs clinicians' diagnoses in real-life patients and usability of web application interface.

For future iterations of deployment, it is recommended to use a more robust infrastructure for running the system in the cloud (e.g. deploying it using containers and Docker/Kubernetes).

In future studies, the generation of synthetic clinical features using the approach used in this paper should be substituted for an alternative of using GANs and VAEs for generating synthetic data, as it will result in more statistically realistic and morphologically coherent distribution of clinical features than Gaussian class-conditional sampling.

5.5 Suggestions for Further Study

The following are suggested fruitful research directions based on this study:

Use of other clinical tools including bone marrow biopsy results, immunophenotyping information and cytogenetic features like Philadelphia chromosome to enrich the fusion architecture in order to build a more sophisticated AI system for haematological disease diagnostics.

Examination of Vision Transformer (ViT) and Hybrid CNN-Transformer models as alternatives to the EfficientNetB3 backbone for extracting features from PBS images, checking whether the attention-based image encoder has additional value in terms of long-range morphological correlations within cells.

Implementation of the federated learning framework for the proposed multi-modal system of leukaemia diagnosis, which will enable training the model using multiple hospitals without transferring personal data. This approach is required due to the inability to build a centralized dataset due to privacy restrictions in African clinical environment.

Deployment Architecture Extension for the inclusion of mobile app interface developed on either React Native or Flutter platform to facilitate point-of-care testing of leukaemia in health centers where desktop web browsing may not be feasible but smartphones are prevalent.

5.6 Conclusion

This research has been able to show that there is a possibility of using a multi-modal deep learning framework that combines the image morphology of Peripheral Blood Smear (PBS) with clinical laboratory information in structure form to be able to obtain extremely high performance when detecting leukaemia by obtaining perfect accuracy, precision, recall, F1-score, and AUC-ROC scores on an out-of-sample test dataset. This work addresses the clinically relevant and empirically proven problem in haematological oncology especially in Nigeria and the greater Sub-Saharan African health care system due to the lack of haematologists and diagnostics infrastructure to make manual PBS examination an unfeasible diagnosis criterion.

Through its implementation of attention mechanisms and transfer learning from powerful pre-trained convolutional networks, the proposed framework contributes to the field of automated haematological diagnosis by transcending the state-of-the-art unimodal image-only model of all comparable systems discussed herein and offering a full-scale deployment of the tool that makes it possible for practitioners to leverage artificial intelligence in their diagnoses. It is shown that the combination of clinical and morphological data under the influence of the attention-based fusion and transfer learning leads to a system which is capable of diagnosing blood diseases on the level equal to and superior to other published specialist-level AI systems.

It is the opinion of the researcher that further development, validation, and contextualization of such diagnostic aids based not only on good machine learning but also on clinical necessity – is one of the most valuable directions for computer science research in the Nigerian and wider African academic environment.

REFERENCES

Acosta, J. N., Falcone, G. J., Rajpurkar, P., & Topol, E. J. (2022). Multimodal biomedical AI. Nature Medicine, 28(9), 1773–1784. https://doi.org/10.1038/s41591-022-01981-2

Adeloye, D., David, R. A., Aderemi, A. V., Iseolorunkanmi, A., Oyedele, O., Ibe, O., Atoloye, A., & Gadanya, M. (2017). An estimate of the incidence of prostate cancer in Africa: A systematic review and meta-analysis. PLOS ONE, 12(5), e0store177113. https://doi.org/10.1371/journal.pone.0177113

Anilkumar, K. K., Bindu, V. R., & Jinesh, K. S. (2021). Classification of leukaemia using ResNet50 deep learning approach. Wireless Personal Communications, 122(3), 2101–2118. https://doi.org/10.1007/s11277-021-08965-4

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056

Gupta, R., & Gupta, P. (2019). ISBI 2019 C-NMC challenge: Classification of normal versus malignant cells. IEEE International Symposium on Biomedical Imaging. https://doi.org/10.1007/978-3-030-31503-0

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4700–4708). https://doi.org/10.1109/CVPR.2017.243

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (NeurIPS), 25 (pp. 1097–1105). https://doi.org/10.1145/3065386

Labati, R. D., Piuri, V., & Scotti, F. (2011). ALL-IDB: The acute lymphoblastic leukaemia image database for image processing. In 18th IEEE International Conference on Image Processing (ICIP) (pp. 2045–2048). https://doi.org/10.1109/ICIP.2011.6115881

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Matek, C., Schwarz, S., Spiekermann, K., & Marr, C. (2019). Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. Nature Machine Intelligence, 1(11),

538–544. https://doi.org/10.1038/s42256-019-0101-9

Ramachandram, D., & Taylor, G. W. (2017). Deep multimodal learning: A survey on recent advances and trends. IEEE Signal Processing Magazine, 34(6), 96–108. https://doi.org/10.1109/MSP.2017.2738401

Rehman, A., Abbas, N., Saba, T., Rahman, S. I. U., Mehmood, Z., & Kolivand, H. (2020). Classification of acute lymphoblastic leukaemia using deep learning. Microscopy Research and Technique, 81(11), 1310–1317. https://doi.org/10.1002/jemt.23139

Shafique, S., & Tehsin, S. (2018). Acute lymphoblastic leukaemia detection and classification of its subtypes using pretrained deep convolutional neural networks. Technology in Cancer Research & Treatment, 17, 1533033818802789. https://doi.org/10.1177/1533033818802789

Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 19, 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging, 35(5), 1285–1298. https://doi.org/10.1109/TMI.2016.2528162

Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, 2023. CA: A Cancer Journal for Clinicians, 73(1), 17–48. https://doi.org/10.3322/caac.21763

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249. https://doi.org/10.3322/caac.21660

Swerdlow, S. H., Campo, E., Harris, N. L., Jaffe, E. S., Pileri, S. A., Stein, H., Thiele, J., & Vardiman, J. W. (Eds.). (2017). WHO classification of tumours of haematopoietic and lymphoid tissues (4th ed.). International Agency for Research on Cancer.

Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (ICML) (pp. 6105–6114). PMLR.

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Tran, T., Vo, T., & Le, T. (2021). Peripheral blood smear analysis using artificial intelligence: A survey. Journal of Digital Imaging, 34(5), 1117–1132. https://doi.org/10.1007/s10278-021-00479-0

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (NeurIPS), 30 (pp. 5998–6008). https://doi.org/10.48550/arXiv.1706.03762

World Health Organization. (2022). WHO classification of tumours: Haematopoietic and lymphoid tumours (5th ed.). International Agency for Research on Cancer.